

# Optimal sequencing of a set of positive numbers with the variance of the sequence's partial sums maximized

Li Wei · Wangdong Qi · Dingxing Chen · Peng Liu · En Yuan

Received: date / Accepted: date

**Abstract** We consider the problem of sequencing a set of positive numbers. We try to find the optimal sequence to maximize the variance of its partial sums. The optimal sequence is shown to have a beautiful structure. It is interesting to note that the symmetric problem which aims at minimizing the variance of the same partial sums is proved to be NP-complete in the literature.

**Keywords** optimal sequence · partial sums · sequencing · variance

## 1 Introduction

This paper considers the problem of sequencing a set of positive numbers to obtain a sequence with the variance of its partial sums maximized (named  $VPS_{\max}$  problem for simplicity). Before giving formal formulation of the problem, we introduce some denotations.

Let  $\mathbf{A} = \{a_k, 1 \leq k \leq n \mid 0 < a_1 < a_2 < \cdots < a_n\}$  be a set of positive numbers sorted in ascending order and  $\mathbf{C} = \pi(\mathbf{A}) = \{c_k, 1 \leq k \leq n\}$  be a sequencing of the elements  $a_1, a_2, \dots, a_n$ . We define the partial sums of sequence  $\mathbf{C}$  as  $\mathbf{S} = \{s_k, 1 \leq k \leq n\}$  with

$$s_k = \sum_{i=1}^k c_i, 1 \leq k \leq n$$

The mean  $\bar{\mathbf{S}}$  and variance  $V(\mathbf{S})$  of  $\mathbf{S}$  are

$$\bar{\mathbf{S}} = \frac{1}{n} \sum_{k=1}^n s_k, V(\mathbf{S}) = \frac{1}{n} \sum_{k=1}^n (s_k - \bar{\mathbf{S}})^2$$

Since  $V(\mathbf{S})$  is also a function of  $\mathbf{C}$ , it can be denoted as  $f(\mathbf{C})$

$$V(\mathbf{S}) = \frac{1}{n} \sum_{k=1}^n \left( \sum_{i=1}^k c_i - \frac{1}{n} \sum_{k=1}^n \sum_{i=1}^k c_i \right)^2 = f(\mathbf{C})$$

The optimal solution to the  $VPS_{\max}$  problem is defined as a sequence  $\mathbf{C}^*$ , formed by sequencing the elements of  $\mathbf{A}$ , satisfying

$$\mathbf{C}^* = \arg \max_{\mathbf{C}} f(\mathbf{C}) = \arg \max_{\mathbf{C}} V(\mathbf{S})$$

---

Li Wei  
PLA University of Science and Technology, Nanjing 210007, China  
E-mail: wlnb@hotmail.com

Wangdong Qi, Dingxing Chen, Peng Liu, En Yuan  
PLA University of Science and Technology, Nanjing 210007, China

That is to say, an optimal solution is a sequence with the variance of its partial sums maximized.

The  $VPS_{\max}$  problem arises from the optimization of the measurement frequencies in a new kind of radio interferometry [9], which is a promising ranging technique in wireless sensor networks. To the best of our knowledge,  $VPS_{\max}$  has not been explored until now.

It should be pointed out that the completion time variance (CTV) problem, which in essence aims at sequencing a set of positive numbers to minimize the variance of the sequence's partial sums, is a symmetric problem of  $VPS_{\max}$ . It has been studied extensively by the operations research community for decades [1]-[8].

In sharp contrast to CTV, which is found to be NP-complete in the literature [5], it is found in this paper that the optimal solution to  $VPS_{\max}$  has a very nice structure.

## 2 Preliminaries

**Definition 1** The mean of the elements in  $\mathbf{S} = \{s_k, 1 \leq k \leq n\}$  ranging from the  $i$ th element to the  $(j-1)$ th element is defined as the  $(i, j)$ -partial mean of  $\mathbf{S}$ , denoted as

$$\mu_{ij}(\mathbf{S}) = \frac{1}{j-i} \sum_{k=i}^{j-1} s_k, 1 \leq i < j \leq n+1$$

Based on the monotonically increasing property of sequence  $\mathbf{S} = \{s_k, 1 \leq k \leq n\}$ , we show that  $\mu_{ij}(\mathbf{S})$  has the following properties:

- (1)  $\bar{\mathbf{S}} = \mu_{1n+1}(\mathbf{S})$ ;
- (2)  $\mu_{ij}(\mathbf{S})$  is a strictly monotonically increasing function of  $i$  and  $j$ ;
- (3) for all  $1 \leq i < j < k \leq n+1$ ,  $\frac{j-i}{k-i} \mu_{ij}(\mathbf{S}) + \frac{k-j}{k-i} \mu_{jk}(\mathbf{S}) = \mu_{ik}(\mathbf{S})$ ;
- (4)  $\mu_{ij}(\mathbf{S}) < \mu_{kl}(\mathbf{S})$  for  $1 \leq i < k < l \leq n+1$  and  $1 \leq i < j < l \leq n+1$ .

To find the sequence that maximize the variance of its partial sums, we try to start from an arbitrary sequencing of  $a_1, a_2, \dots, a_n$  and follow a path of favorable transforms that would eventually lead to the optimal one. We begin with the simplest transform from one sequence to another by interchanging positions of only two elements.

**Definition 2** The transform from the sequence  $\mathbf{C} = \{c_1, \dots, c_{i-1}, c_i, c_{i+1}, \dots, c_{j-1}, c_j, c_{j+1}, \dots, c_n\}$  to  $\mathbf{C}' = \{c_1, \dots, c_{i-1}, c_j, c_{i+1}, \dots, c_{j-1}, c_i, c_{j+1}, \dots, c_n\}$  by interchanging the  $i$ th element and the  $j$ th element is called an  $(i, j)$ -interchange of  $\mathbf{C}$ .

An  $(i, j)$ -interchange from  $\mathbf{C}$  to  $\mathbf{C}'$  is a *favorable transform* and  $\mathbf{C}'$  is called a *better* sequence if  $f(\mathbf{C}') > f(\mathbf{C})$ .

Next, we give a criterion to determine whether a given  $(i, j)$ -interchange is a *favorable* transform.

Let  $\mathbf{C}'$  denotes the transform of  $\mathbf{C} = \{c_k, 1 \leq k \leq n\}$  via  $(i, j)$ -interchange and  $\delta = c_j - c_i$ . We get

$$s'_k = \begin{cases} s_k + \delta & i \leq k < j \\ s_k & \text{others} \end{cases}$$

and

$$\bar{\mathbf{S}}' = \bar{\mathbf{S}} + \frac{(j-i)\delta}{n}$$

Since the objective function can be simplified as

$$f(\mathbf{C}) = V(\mathbf{S}) = \frac{1}{n} \sum_{k=1}^n s_k^2 - \bar{\mathbf{S}}^2 \quad (1)$$

It follows that

$$\begin{aligned}
 f(\mathbf{C}') - f(\mathbf{C}) &= \left( \frac{1}{n} \sum_{k=1}^n s_k'^2 - \bar{\mathbf{S}}^2 \right) - \left( \frac{1}{n} \sum_{k=1}^n s_k^2 - \bar{\mathbf{S}}^2 \right) \\
 &= \left( \frac{1}{n} \sum_{k=i}^{j-1} ((s_k + \delta)^2 - s_k^2) \right) - \left( \left( \bar{\mathbf{S}} + \frac{(j-i)\delta}{n} \right)^2 - \bar{\mathbf{S}}^2 \right) \\
 &= \left( \frac{2\delta}{n} \sum_{k=i}^{j-1} s_k + \frac{(j-i)}{n} \delta^2 \right) - \left( \frac{2(j-i)\delta}{n} \bar{\mathbf{S}} + \frac{(j-i)^2 \delta^2}{n^2} \right) \\
 &= \frac{2(j-i)\delta}{n} \left( \frac{1}{j-i} \sum_{k=i}^{j-1} s_k - \bar{\mathbf{S}} + \frac{\delta}{2} \left( 1 - \frac{(j-i)}{n} \right) \right)
 \end{aligned}$$

In latter parts of this paper, we denote the difference  $f(\mathbf{C}') - f(\mathbf{C})$  as  $\Delta f(i, j, \mathbf{C}, \mathbf{C}')$ . Then

$$\Delta f(i, j, \mathbf{C}, \mathbf{C}') = \frac{2(j-i)\delta}{n} \left( \mu_{ij}(\mathbf{S}) - \bar{\mathbf{S}} + \frac{\delta}{2} \left( 1 - \frac{(j-i)}{n} \right) \right) \quad (2)$$

Let  $\varphi(\delta) = D_1 \delta^2 + D_2 \delta = \Delta f(i, j, \mathbf{C}, \mathbf{C}')$ , where  $D_1 = \frac{j-i}{n} \left( 1 - \frac{j-i}{n} \right) > 0$  and  $D_2 = \frac{2(j-i)}{n} (\mu_{ij}(\mathbf{S}) - \bar{\mathbf{S}})$ . Based on the property of quadratic equation, we have criterions for *favorable (i, j)-interchange*.

- Claim 1: if  $D_2 > 0$  (i.e.  $\mu_{ij}(\mathbf{S}) > \bar{\mathbf{S}}$ ), then the *(i, j)-interchange* is *favorable* when  $\delta > 0$  or  $\delta < -D_2/D_1$ .
- Claim 2: if  $D_2 \leq 0$  (i.e.  $\mu_{ij}(\mathbf{S}) \leq \bar{\mathbf{S}}$ ), then the *(i, j)-interchange* is *favorable* when  $\delta < 0$  or  $\delta > -D_2/D_1$ .

We then have the following results.

**Proposition 1** *If  $c_1 \neq a_1$ , then sequence  $\mathbf{C}$  is not the optimal solution.*

*Proof* Clearly,  $\mu_{1j}(\mathbf{S}) < \mu_{1n+1}(\mathbf{S}) = \bar{\mathbf{S}}$ , for any integer  $j \in [2, n]$ . Therefore, if  $a_1 = c_j$ , we will obtain a *better* sequence than  $\mathbf{C}$  via *(1, j)-interchange* according to claim 2.  $\square$

**Proposition 2** *If  $c_n = a_n$  with  $n > 3$ , then sequence  $\mathbf{C}$  is not the optimal solution.*

*Proof* Note that

$$\mu_{n-1n}(\mathbf{S}) = \sum_{k=1}^{n-1} c_k, \bar{\mathbf{S}} = \sum_{k=1}^n \frac{n-k+1}{n} c_k, \delta = c_n - c_{n-1} > 0$$

and

$$\begin{aligned}
 \Delta f(n-1, n, \mathbf{C}, \mathbf{C}') &= \frac{2\delta}{n} \left( \mu_{n-1n}(\mathbf{S}) - \bar{\mathbf{S}} + \frac{\delta}{2} \left( \frac{n-1}{n} \right) \right) \\
 &= \frac{2\delta}{n} \left( \sum_{k=1}^{n-1} \frac{k-1}{n} c_k - \frac{c_n}{n} + \frac{(n-1)(c_n - c_{n-1})}{2n} \right) \\
 &= \frac{2\delta}{n} \left( \sum_{k=1}^{n-2} \frac{k-1}{n} c_k + \frac{n-2}{n} c_{n-1} - \frac{c_n}{n} + \frac{(n-1)(c_n - c_{n-1})}{2n} \right) \\
 &= \frac{2(c_n - c_{n-1})}{n} \left( \sum_{k=1}^{n-2} \frac{k-1}{n} c_k + \frac{(n-3)(c_n + c_{n-1})}{2n} \right)
 \end{aligned}$$

The expression is always positive for  $n > 3$ , i.e.  $\Delta f(n-1, n, \mathbf{C}, \mathbf{C}') > 0$ . So, *(n-1, n)-interchange* will obtain a *better* sequence than  $\mathbf{C}$ .  $\square$

**Proposition 3** *If sequence  $\mathbf{C}$  is the optimal solution with  $c_k = a_n, 1 < k < n$ , then*

- (1) for any  $i, j$  with  $1 \leq i < j \leq k$ ,  $\mu_{ij}(\mathbf{S}) < \bar{S}$ ;  
 (2) for any  $i, j$  with  $k \leq i < j \leq n$ ,  $\mu_{ij}(\mathbf{S}) > \bar{S}$ .

*Proof* (1) Otherwise, there exist  $i, j$  such that  $1 \leq i < j \leq k$  and  $\mu_{ij}(\mathbf{S}) \geq \bar{S}$ , then

$$\mu_{ik}(\mathbf{S}) \geq \mu_{ij}(\mathbf{S}) \geq \bar{S}$$

A better sequence will be created when exchanging the  $i$ th and the  $k$ th element of  $\mathbf{C}$  due to  $\delta = c_k - c_i > 0$ . This conflicts with the optimality of  $\mathbf{C}$ .

A similar proof can be applied to (2).  $\square$

**Definition 3** A sequence is called a  $\wedge$ -shaped sequence when the elements before the largest one are sorted in an ascending order, while the elements after the largest one are sorted in a descending order.

**Proposition 4** The optimal sequence is  $\wedge$ -Shaped. In other words, if sequence  $\mathbf{C}$  is the optimal solution with  $c_k = a_n$ ,  $1 < k < n$ , then

- (1) for any  $i, j$  with  $1 \leq i < j \leq k$ ,  $c_i < c_j$ ;  
 (2) for any  $i, j$  with  $k \leq i < j \leq n$ ,  $c_i > c_j$ .

*Proof* It is a direct conclusion of Proposition 3.  $\square$

**Remark.** There exists similar property for the CTV problem. Eilon and Chowdhury [4] proved that the optimal solutions to the CTV minimization problem should be V-shaped, meaning that the elements before the smallest one are sorted in a decreasing order, while the elements after the smallest one are sorted in an ascending order.

### 3 Main results

**Definition 4** For a sequence  $\mathbf{C} = \{c_k, 1 \leq k \leq n\}$ , do  $(k, n+2-k)$ -interchange for all  $2 \leq k \leq u$ , and we can obtain the dual sequence  $\mathbf{C}^d$  of  $\mathbf{C}$ , where  $u = \lceil n/2 \rceil$ ,  $\lceil \cdot \rceil$  denotes the ceiling function.

**Lemma 1** For sequence  $\mathbf{C} = \{c_1, c_2, c_3, \dots, c_{n-1}, c_n\}$  and its dual sequence  $\mathbf{C}^d = \{c_1, c_n, c_{n-1}, \dots, c_3, c_2\}$ , the following identity holds.

$$f(\mathbf{C}) = f(\mathbf{C}^d)$$

*Proof* See [1], we also provide another proof of Lemma 1 in the appendix.  $\square$

**Lemma 2** There exist at least two kinds of optimal sequences, one is in the form of  $c_1 = a_1, c_2 = a_2$ , and the other is in the form of  $c_1 = a_1, c_n = a_2$ .

*Proof* We consider the position of  $a_n$  in the optimal sequence first. Since  $c_1 = a_1$  and  $c_n \neq a_n$ , by Proposition 4, we have

- (1) case  $c_2 = a_n$ : then  $c_n = a_2$ ;  
 (2) case  $c_k = a_n, 2 < k < n$ : then  $c_n = a_2$  or  $c_2 = a_2$ ;

This means either  $c_n = a_2$  or  $c_2 = a_2$ . When  $c_n = a_2$ ,  $c_2 = a_2$  is also the optimal position for  $a_2$  by Lemma 1 and vice versa. Therefore, both  $c_n = a_2$  and  $c_2 = a_2$  are the optimal position for  $a_2$ .  $\square$

**Remark.** Note that the optimal sequence of the CTV problem is in the form of  $c_1 = a_n, c_2 = a_{n-1}$  or  $c_1 = a_n, c_n = a_{n-1}$  [2] [3]. Similarity appears once again. The major difference will be shown below.

**Definition 5** For a sequence  $\mathbf{C} = \{c_k, 1 \leq k \leq n\}$ , if the following two conditions do not hold,

- (1)  $c_k < c_{n+2-k}$ , for all  $2 \leq k \leq u$ ;

(2)  $c_k > c_{n+2-k}$ , for all  $2 \leq k \leq u$ .

then do  $(k, n+2-k)$ -interchange for each  $k$  satisfying  $c_k < c_{n+2-k}$ ,  $2 \leq k \leq u$  until all  $c_k > c_{n+2-k}$ ,  $2 \leq k \leq u$ . This is called the *sum-‘n+2’ transform* of  $\mathbf{C}$ .

**Theorem 1** A sum-‘n+2’ transform always results in a better sequence.

*Proof* Note that

$$\mu_{ij}(\mathbf{S}) = \sum_{k=1}^{i-1} c_k + \sum_{k=i}^{j-1} \frac{j-k}{j-i} c_k$$

We begin by rewriting (2) as:

$$\begin{aligned} \Delta f(i, j, \mathbf{C}, \mathbf{C}') &= \frac{2(j-i)(c_j - c_i)}{n} \left( \sum_{k=1}^{i-1} c_k + \sum_{k=i}^{j-1} \frac{j-k}{j-i} c_k - \sum_{k=1}^n \frac{n+1-k}{n} c_k + \frac{c_j - c_i}{2} \left( 1 - \frac{j-i}{n} \right) \right) \\ &= \frac{2(j-i)(c_j - c_i)}{n^2} \left( \sum_{k=1}^{i-1} (k-1)c_k + \frac{1}{j-i} \sum_{k=i+1}^{j-1} ((j-i-n)k - j + i + ni)c_k \right. \\ &\quad \left. - \sum_{k=j+1}^n (n+1-k)c_k + (i+j-n-2)(c_j + c_i)/2 \right) \end{aligned} \quad (3)$$

If  $i+j = n+2$ , we obtain

$$\begin{aligned} \Delta f(i, n+2-i, \mathbf{C}, \mathbf{C}') &= \frac{2(n+2-2i)(c_{n+2-i} - c_i)}{n^2} \left( \sum_{k=2}^{i-1} (k-1)(c_k - c_{n+2-k}) + \frac{i-1}{n+2-2i} \sum_{k=i+1}^{n+1-i} (n+2-2k)c_k \right) \\ &= \frac{2(n+2-2i)(c_{n+2-i} - c_i)}{n^2} \left( \sum_{k=2}^{i-1} (k-1)(c_k - c_{n+2-k}) \right. \\ &\quad \left. + \frac{i-1}{n+2-2i} \sum_{k=i+1}^u (n+2-2k)(c_k - c_{n+2-k}) \right) \\ &= \frac{2(c_{n+2-i} - c_i)}{n^2} \left( (n+2-2i) \sum_{k=2}^{i-1} (k-1)(c_k - c_{n+2-k}) \right. \\ &\quad \left. + (i-1) \sum_{k=i+1}^u (n+2-2k)(c_k - c_{n+2-k}) \right) \end{aligned} \quad (4)$$

where  $u = \lceil n/2 \rceil$ .

Without loss of generality, Let  $I = \{i_1, i_2, \dots, i_p \mid 2 \leq i_1 < i_2 < \dots < i_p \leq u\}$  denote the aggregate set of the indexes  $k$  in  $\mathbf{C} = \{c_k, k = 1 \dots n\}$  satisfying  $c_{i_q} < c_{n+2-i_q}$ ,  $q = 1, 2, \dots, p$ .

We define  $\mathbf{C}^{(m)} = \{c_k^{(m)}, k = 1 \dots n, m \leq p\}$  as the transform of sequence  $\mathbf{C}$  via  $(i_1, n+2-i_1), (i_2, n+2-i_2) \dots (i_m, n+2-i_m)$ -interchange, which is also denoted as  $\mathbf{C}^{(0)}$  for notational convenience. By definition,  $\mathbf{C}^{(m)}$  may also be obtained from  $\mathbf{C}^{(m-1)}$  by  $(i_m, n+2-i_m)$ -interchange. The element  $c_k^{(m)}$  is given by

$$\begin{aligned} c_k^{(m)} &= \begin{cases} c_{n+2-k}^{(m-1)} & k = i_m \\ c_{n+2-k}^{(m-1)} & k = n+2-i_m \\ c_k^{(m-1)} & \text{others} \end{cases} \\ &= \begin{cases} c_{n+2-k} & k = i_q, 1 \leq q \leq m \\ c_{n+2-k} & k = n+2-i_q, 1 \leq q \leq m \\ c_k & \text{others} \end{cases} \end{aligned} \quad (5)$$

We then have

$$\begin{aligned} f(\mathbf{C}^{(m)}) - f(\mathbf{C}^{(m-1)}) &= \Delta f(i_m, n+2-i_m, \mathbf{C}^{(m-1)}, \mathbf{C}^{(m)}) \\ &= \frac{2}{n^2} (c_{n+2-i_m}^{(m-1)} - c_{i_m}^{(m-1)}) \left( (n+2-2i_m) \sum_{k=2}^{i_m-1} (k-1) (c_k^{(m-1)} - c_{n+2-k}^{(m-1)}) \right. \\ &\quad \left. + (i_m-1) \sum_{k=i_m+1}^u (n+2-2k) (c_k^{(m-1)} - c_{n+2-k}^{(m-1)}) \right) \end{aligned}$$

It follows from (5) that

$$\begin{aligned} \Delta f(i_m, n+2-i_m, \mathbf{C}^{(m-1)}, \mathbf{C}^{(m)}) &= \frac{2}{n^2} (c_{n+2-i_m} - c_{i_m}) \left[ (n+2-2i_m) \right. \\ &\quad \left( \sum_{k \in I, k < i_m} (k-1) (c_k^{(m-1)} - c_{n+2-k}^{(m-1)}) + \sum_{k \notin I, 2 \leq k < i_m} (k-1) (c_k^{(m-1)} - c_{n+2-k}^{(m-1)}) \right) \\ &\quad \left. + (i_m-1) \sum_{k=i_m+1}^u (n+2-2k) (c_k - c_{n+2-k}) \right] \\ &= \frac{2}{n^2} (c_{n+2-i_m} - c_{i_m}) \left[ (n+2-2i_m) \left( \sum_{k \in I, k < i_m} (k-1) (c_{n+2-k} - c_k) \right. \right. \\ &\quad \left. \left. + \sum_{k \notin I, 2 \leq k < i_m} (k-1) (c_k - c_{n+2-k}) \right) + (i_m-1) \left( \sum_{k \in I, k > i_m} (n+2-2k) (c_k - c_{n+2-k}) \right. \right. \\ &\quad \left. \left. + \sum_{k \notin I, i_m < k \leq u} (n+2-2k) (c_k - c_{n+2-k}) \right) \right] \quad (6) \end{aligned}$$

Let

$$\begin{aligned} Z(i_m, n+2-i_m, \mathbf{C}^{(m-1)}, \mathbf{C}^{(m)}) &= \frac{2}{n^2} (c_{n+2-i_m} - c_{i_m}) \\ &\quad \left( (n+2-2i_m) \sum_{k \in I, k < i_m} (k-1) (c_{n+2-k} - c_k) + (i_m-1) \sum_{k \in I, k > i_m} (n+2-2k) (c_k - c_{n+2-k}) \right) \end{aligned}$$

and

$$\begin{aligned} R(i_m, n+2-i_m, \mathbf{C}^{(m-1)}, \mathbf{C}^{(m)}) &= \frac{2}{n^2} (c_{n+2-i_m} - c_{i_m}) \\ &\quad \left( (n+2-2i_m) \sum_{k \notin I, 2 \leq k < i_m} (k-1) (c_k - c_{n+2-k}) + (i_m-1) \sum_{k \notin I, i_m < k \leq u} (n+2-2k) (c_k - c_{n+2-k}) \right) \end{aligned}$$

Note that, using the definitions, the whole increment of  $f(\mathbf{C}^{(0)})$  after the *sum-‘n+2’ transform* from  $\mathbf{C}^{(0)}$  to  $\mathbf{C}^{(p)}$  (via  $(i_1, n+2-i_1), (i_2, n+2-i_2) \dots (i_p, n+2-i_p)$ -interchange) is given by

$$\begin{aligned} f(\mathbf{C}^{(p)}) - f(\mathbf{C}^{(0)}) &= \sum_{m=1}^p (f(\mathbf{C}^{(m)}) - f(\mathbf{C}^{(m-1)})) \\ &= \sum_{m=1}^p [Z(i_m, n+2-i_m, \mathbf{C}^{(m-1)}, \mathbf{C}^{(m)}) + R(i_m, n+2-i_m, \mathbf{C}^{(m-1)}, \mathbf{C}^{(m)})] \quad (7) \end{aligned}$$

We then prove the following identity

$$\begin{aligned}
\sum_{m=1}^p Z(i_m, n+2-i_m, \mathbf{C}^{(m-1)}, \mathbf{C}^{(m)}) &= \frac{2}{n^2} \left( \sum_{m=1}^p (n+2-2i_m)(c_{n+2-i_m} - c_{i_m}) \sum_{k \in I, k < i_m} (k-1)(c_{n+2-k} - c_k) \right. \\
&\quad \left. - \sum_{m=1}^p (i_m-1)(c_{n+2-i_m} - c_{i_m}) \sum_{k \in I, k > i_m} (n+2-2k)(c_{n+2-k} - c_k) \right) \\
&= \frac{2}{n^2} \left( \sum_{m=2}^p \sum_{l=1}^{m-1} (n+2-2i_m)(c_{n+2-i_m} - c_{i_m})(i_l-1)(c_{n+2-i_l} - c_{i_l}) \right. \\
&\quad \left. - \sum_{m=1}^{p-1} \sum_{l=m+1}^p (i_m-1)(c_{n+2-i_m} - c_{i_m})(n+2-2i_l)(c_{n+2-i_l} - c_{i_l}) \right) \\
&= \frac{2}{n^2} \left( \sum_{l=1}^{p-1} \sum_{m=l+1}^p (i_l-1)(c_{n+2-i_l} - c_{i_l})(n+2-2i_m)(c_{n+2-i_m} - c_{i_m}) \right. \\
&\quad \left. - \sum_{m=1}^{p-1} \sum_{l=m+1}^p (i_m-1)(c_{n+2-i_m} - c_{i_m})(n+2-2i_l)(c_{n+2-i_l} - c_{i_l}) \right) \\
&= 0
\end{aligned} \tag{8}$$

Therefore, the residual part

$$\begin{aligned}
f(\mathbf{C}^{(p)}) - f(\mathbf{C}^{(0)}) &= \sum_{m=1}^p R(i_m, n+2-i_m, \mathbf{C}^{(m-1)}, \mathbf{C}^{(m)}) \\
&= \frac{2}{n^2} \left( \sum_{m=1}^p (n+2-2i_m)(c_{n+2-i_m} - c_{i_m}) \sum_{k \notin I, 2 \leq k < i_m} (k-1)(c_k - c_{n+2-k}) \right. \\
&\quad \left. + \sum_{m=1}^p (i_m-1)(c_{n+2-i_m} - c_{i_m}) \sum_{k \notin I, i_m < k \leq u} (n+2-2k)(c_k - c_{n+2-k}) \right)
\end{aligned} \tag{9}$$

The set  $K = \{k \notin I, 2 \leq k < i_m, 1 \leq m \leq p\} \cup \{k \notin I, i_m < k \leq u, 1 \leq m \leq p\}$  is not empty by definition. Otherwise, we will get a contradiction with condition (1) of definition 5.

Considering the inequality  $(c_{n+2-i_m} - c_{i_m})(c_k - c_{n+2-k}) > 0$  when  $k \notin I$  and  $(n+2-2k) > 0$  when  $k \leq u$ , we finally obtain

$$f(\mathbf{C}^{(p)}) - f(\mathbf{C}^{(0)}) > 0$$

Thus the proof is complete.  $\square$

**Example 1.** Let us regard a sequence  $C^{(0)} = [1, 6, 2, 3, 4, 8, 7, 5]$ , then  $C^{(p)} = C^{(2)} = [1, 6, 7, 8, 4, 3, 2, 5]$  is its *sum- $'n+2'$  transform*. Note that the transform consists of  $(3, 7)$ -interchange and  $(4, 6)$ -interchange. So we get the set  $I = \{i_1, i_2, \dots, i_p \mid 2 \leq i_1 < i_2 < \dots < i_p \leq u, c_{i_q} < c_{n+2-i_q}, q = 1, 2, \dots, p\} = \{i_1 = 3, i_2 = 4\}$ .

From (9), we know

$$\begin{aligned}
f(\mathbf{C}^{(p)}) - f(\mathbf{C}^{(0)}) &= \sum_{m=1}^2 R(i_m, n+2-i_m, \mathbf{C}^{(m-1)}, \mathbf{C}^{(m)}) \\
&= \frac{2}{n^2} (c_7 - c_3) \left[ (n+2-2i_1) \sum_{k \notin I, 2 \leq k < i_1} (k-1)(c_k - c_{n+2-k}) + (i_1-1) \sum_{k \notin I, i_1 < k \leq u} (n+2-2k)(c_k - c_{n+2-k}) \right] \\
&\quad + \frac{2}{n^2} (c_6 - c_4) \left[ (n+2-2i_2) \sum_{k \notin I, 2 \leq k < i_2} (k-1)(c_k - c_{n+2-k}) + (i_2-1) \sum_{k \notin I, i_2 < k \leq u} (n+2-2k)(c_k - c_{n+2-k}) \right] \\
&= \frac{2}{n^2} (c_7 - c_3) [(n+2-2i_1)(2-1)(c_2 - c_8)] + \frac{2}{n^2} (c_6 - c_4) [(n+2-2i_2)(2-1)(c_2 - c_8)] \\
&= 0.9375
\end{aligned}$$

On the other hand, we know  $f(C^{(0)}) = 131.5$  and  $f(C^{(2)}) = 132.4375$ . It follows that

$$f(C^{(2)}) - f(C^{(0)}) = 0.9375$$

This verifies the identity (9) as well as Theorem 1.

**Definition 6** For sequence  $\mathbf{C} = \{c_k, 1 \leq k \leq n\}$ , do  $(k, n+1-k)$ -interchange for each  $k$  satisfying  $c_k > c_{n+1-k}$ ,  $2 \leq k \leq u'$  until all  $c_k < c_{n+1-k}$ ,  $2 \leq k \leq u'$ . where  $u' = \lfloor n/2 \rfloor$ ,  $\lfloor \cdot \rfloor$  denotes the floor function. This is called the *sum- $'n+1'$  transform* of  $\mathbf{C}$ .

**Theorem 2** A *sum- $'n+1'$  transform* always results in a better sequence.

*Proof* Without loss of generality, Let  $I' = \{i_1, i_2, \dots, i_{p'} \mid 2 \leq i_1 < i_2 < \dots < i_{p'} \leq u'\}$  denote the aggregate set of the indexes  $k$  in  $\mathbf{C} = \{c_k, 1 \leq k \leq n\}$  satisfying  $c_{i_q} > c_{n+1-i_q}$ ,  $q = 1, 2, \dots, p'$ ,  $u' = \lfloor n/2 \rfloor$ .

If  $i + j = n + 1$ , from (3), we obtain

$$\begin{aligned} \Delta f(i, n+1-i, \mathbf{C}, \mathbf{C}') &= \frac{2}{n^2} (n+1-2i)(c_{n+1-i} - c_i) \left[ - \sum_{k=n+2-i}^n (n+1-k)c_k \right. \\ &\quad \left. - (c_{n+1-i} + c_i)/2 + \sum_{k=2}^{i-1} (k-1)c_k + \frac{1}{n+1-2i} \sum_{k=i+1}^{n-i} ((1-2i)k - n - 1 + 2i + ni)c_k \right] \\ &= \frac{2}{n^2} (n+1-2i)(c_{n+1-i} - c_i) \left[ - \sum_{k=1}^{i-1} c_{n+1-k} - \frac{1}{n+1-2i} \sum_{k=i}^{n-i} (k-i)c_k \right. \\ &\quad \left. - (c_{n+1-i} + c_i)/2 + \sum_{k=2}^{i-1} (k-1)(c_k - c_{n+1-k}) \right. \\ &\quad \left. + \frac{i-1}{n+1-2i} \sum_{k=i+1}^{u'} (n+1-2k)(c_k - c_{n+1-k}) \right] \quad (10) \end{aligned}$$

We define  $\tilde{\mathbf{C}}^{(m)} = \{\tilde{c}_k^{(m)}, k = 1 \dots n, m \leq p'\}$  as the transform of sequence  $\mathbf{C}$  via  $(i_1, n+1-i_1), (i_2, n+1-i_2) \dots (i_m, n+1-i_m)$ -interchange, which is also denoted as  $\tilde{\mathbf{C}}^{(0)}$  for notational convenience. By definition,  $\tilde{\mathbf{C}}^{(m)}$  may also be obtained from  $\tilde{\mathbf{C}}^{(m-1)}$  by  $(i_m, n+1-i_m)$ -interchange. The element  $\tilde{c}_k^{(m)}$  is given by

$$\begin{aligned} \tilde{c}_k^{(m)} &= \begin{cases} \tilde{c}_{n+1-k}^{(m-1)} & k = i_m \\ \tilde{c}_{n+1-k}^{(m-1)} & k = n+1-i_m \\ \tilde{c}_k^{(m-1)} & \text{others} \end{cases} \\ &= \begin{cases} c_{n+1-k} & k = i_q, 1 \leq q \leq m \\ c_{n+1-k} & k = n+1-i_q, 1 \leq q \leq m \\ c_k & \text{others} \end{cases} \quad (11) \end{aligned}$$



Then

$$\begin{aligned}
f(\tilde{\mathbf{C}}^{(m)}) - f(\tilde{\mathbf{C}}^{(m-1)}) &= \Delta f(i_m, n+1-i_m, \tilde{\mathbf{C}}^{(m-1)}, \tilde{\mathbf{C}}^{(m)}) \\
&= \frac{2}{n^2} (n+1-2i_m) \left( \tilde{c}_{n+1-i_m}^{(m-1)} - \tilde{c}_{i_m}^{(m-1)} \right) \left[ - \sum_{k=1}^{i_m-1} \tilde{c}_{n+1-k}^{(m-1)} \right. \\
&\quad - \frac{1}{n+1-2i_m} \sum_{k=i_m}^{n-i_m} (k-i_m) \tilde{c}_k^{(m-1)} + \sum_{k=2}^{i_m-1} (k-1) \left( \tilde{c}_k^{(m-1)} - \tilde{c}_{n+1-k}^{(m-1)} \right) \\
&\quad \left. - \left( \tilde{c}_{n+1-i_m}^{(m-1)} + \tilde{c}_{i_m}^{(m-1)} \right) / 2 + \frac{i_m-1}{n+1-2i_m} \sum_{k=i_m+1}^{u'} (n+1-2k) \left( \tilde{c}_k^{(m-1)} - \tilde{c}_{n+1-k}^{(m-1)} \right) \right] \\
&= \frac{2}{n^2} (n+1-2i_m) (c_{n+1-i_m} - c_{i_m}) \left[ - \sum_{k=1}^{i_m-1} \tilde{c}_{n+1-k}^{(m-1)} - \frac{1}{n+1-2i_m} \sum_{k=i_m}^{n-i_m} (k-i_m) \tilde{c}_k^{(m-1)} \right. \\
&\quad - \left( \tilde{c}_{n+1-i_m}^{(m-1)} + \tilde{c}_{i_m}^{(m-1)} \right) / 2 + \sum_{k=2}^{i_m-1} (k-1) \left( \tilde{c}_k^{(m-1)} - \tilde{c}_{n+1-k}^{(m-1)} \right) \\
&\quad \left. + \frac{i_m-1}{n+1-2i_m} \sum_{k=i_m+1}^{u'} (n+1-2k) (c_k - c_{n+1-k}) \right] \tag{12}
\end{aligned}$$

Note that for  $i \in I', k \notin I'$ , we have

$$\begin{aligned}
(c_{n+1-i} - c_i) &< 0, \\
(c_{n+1-i} - c_i)(c_k - c_{n+1-k}) &> 0
\end{aligned}$$

It follows that

$$\begin{aligned}
\Delta f(i_m, n+1-i_m, \tilde{\mathbf{C}}^{(m-1)}, \tilde{\mathbf{C}}^{(m)}) &> \frac{2}{n^2} (c_{n+1-i_m} - c_{i_m}) \left[ (n+1-2i_m) \sum_{k=2}^{i_m-1} (k-1) \left( \tilde{c}_k^{(m-1)} - \tilde{c}_{n+1-k}^{(m-1)} \right) \right. \\
&\quad \left. + (i_m-1) \sum_{k=i_m+1}^{u'} (n+1-2k) (c_k - c_{n+1-k}) \right] \\
&= \frac{2}{n^2} (c_{n+1-i_m} - c_{i_m}) \left[ (n+1-2i_m) \left( \sum_{k \in I', k < i_m} (k-1) \left( \tilde{c}_k^{(m-1)} - \tilde{c}_{n+1-k}^{(m-1)} \right) \right. \right. \\
&\quad \left. \left. + \sum_{k \notin I', 2 \leq k < i_m} (k-1) (c_k - c_{n+1-k}) \right) + (i_m-1) \right. \\
&\quad \left. \left( \sum_{k \notin I', i_m < k \leq u'} (n+1-2k) (c_k - c_{n+1-k}) + \sum_{k \in I', k > i_m} (n+1-2k) (c_k - c_{n+1-k}) \right) \right] \\
&> \frac{2}{n^2} (c_{n+1-i_m} - c_{i_m}) \left[ (n+1-2i_m) \sum_{k \in I', k < i_m} (k-1) \left( \tilde{c}_k^{(m-1)} - \tilde{c}_{n+1-k}^{(m-1)} \right) \right. \\
&\quad \left. + (i_m-1) \sum_{k \in I', k > i_m} (n+1-2k) (c_k - c_{n+1-k}) \right] \\
&= \frac{2}{n^2} (c_{n+1-i_m} - c_{i_m}) \left[ (n+1-2i_m) \sum_{k \in I', k < i_m} (k-1) (c_{n+1-k} - c_k) \right. \\
&\quad \left. + (i_m-1) \sum_{k \in I', k > i_m} (n+1-2k) (c_k - c_{n+1-k}) \right] \tag{13}
\end{aligned}$$

Following derivations similar to (8) yield

$$\sum_{m=1}^{p'} (c_{n+1-i_m} - c_{i_m}) \left[ (n+1-2i_m) \sum_{k \in I', k < i_m} (k-1)(c_{n+1-k} - c_k) + (i_m-1) \sum_{k \in I', k > i_m} (n+1-2k)(c_k - c_{n+1-k}) \right] = 0$$

So we may conclude that

$$f(\tilde{\mathbf{C}}^{(p')}) - f(\tilde{\mathbf{C}}^{(0)}) = \sum_{m=1}^{p'} \Delta f(i_m, n+1-i_m, \tilde{\mathbf{C}}^{(m-1)}, \tilde{\mathbf{C}}^{(m)}) > 0$$

Where  $\tilde{\mathbf{C}}^{(p')}$  is the *sum- $'n+1'$  transform* of  $\tilde{\mathbf{C}}^{(0)}$ , the conclusion follows.  $\square$

**Theorem 3** *If the optimal sequence  $\mathbf{C}^*$  is in the form of  $c_1^* = a_1$  and  $c_n^* = a_2$ , then the following hold.*

- (1) *Case  $n$  is even or  $u = u'$  :  $c_{n+2-k}^* < c_k^* < c_{n+1-k}^*$  for  $2 \leq k \leq u$ ;*
- (2) *Case  $n$  is odd or  $u = u' + 1$  :  $c_k^* < c_{n+1-k}^*$  for  $2 \leq k \leq u'$  and  $c_{n+2-k}^* < c_k^*$  for  $2 \leq k \leq u$ .*

*Proof* We will first prove  $c_k^* > c_{n+2-k}^*$  by contradiction. It is obvious that  $c_2^* > c_n^*$ , then the condition (1) of definition 5 doesn't hold. If the condition (2) of definition 5 is also not satisfied, that is, there must exist a set  $I = \{i_1, i_2, \dots, i_p \mid 2 \leq i_1 < i_2 < \dots < i_p \leq u, c_{i_q}^* < c_{n+2-i_q}^*, q = 1, 2, \dots, p\}$ . Now we will get a *better* sequence by applying *sum- $'n+2'$  transform*. A contradiction.

This implies that the condition (2) must hold. In other words, we obtain  $c_k^* > c_{n+2-k}^*$  for  $2 \leq k \leq u$ .

Similarly, we have  $c_k^* < c_{n+1-k}^*$  for  $2 \leq k \leq u'$ .

This completes the proof.  $\square$

**Corollary 1** *The optimal sequences are*

$$\mathbf{C}^* = \begin{cases} \{a_1, a_3, a_5, a_7, \dots, a_n, a_{n-1}, a_{n-3}, \dots, a_6, a_4, a_2\} & n \text{ is odd} \\ \{a_1, a_3, a_5, a_7, \dots, a_{n-1}, a_n, a_{n-2}, \dots, a_6, a_4, a_2\} & n \text{ is even} \end{cases}$$

and

$$\mathbf{C}^{*d} = \begin{cases} \{a_1, a_2, a_4, a_6, \dots, a_{n-1}, a_n, a_{n-2}, \dots, a_7, a_5, a_3\} & n \text{ is odd} \\ \{a_1, a_2, a_4, a_6, \dots, a_n, a_{n-1}, a_{n-3}, \dots, a_7, a_5, a_3\} & n \text{ is even} \end{cases}$$

*Proof* By Theorem 3 and Lemma 2, we readily obtain the single optimal solution  $\mathbf{C}^*$  in the form of  $c_1^* = a_1$  and  $c_n^* = a_2$ ,

$$\mathbf{C}^* = \begin{cases} \{a_1, a_3, a_5, a_7, \dots, a_n, a_{n-1}, a_{n-3}, \dots, a_6, a_4, a_2\} & n \text{ is odd} \\ \{a_1, a_3, a_5, a_7, \dots, a_{n-1}, a_n, a_{n-2}, \dots, a_6, a_4, a_2\} & n \text{ is even} \end{cases}$$

Then we immediately obtain another optimal solution of the form  $c_1^{*d} = a_1$  and  $c_2^{*d} = a_2$  by Lemma 1

$$\mathbf{C}^{*d} = \begin{cases} \{a_1, a_2, a_4, a_6, \dots, a_{n-1}, a_n, a_{n-2}, \dots, a_7, a_5, a_3\} & n \text{ is odd} \\ \{a_1, a_2, a_4, a_6, \dots, a_n, a_{n-1}, a_{n-3}, \dots, a_7, a_5, a_3\} & n \text{ is even} \end{cases}$$

We claim that the optimal solution of the form  $c_1^{*d} = a_1$  and  $c_2^{*d} = a_2$  is also unique. Otherwise, we may get more than one optimal solution of the form  $c_1^* = a_1$  and  $c_n^* = a_2$  according to Lemma 1, contradiction.

Hence we obtain all the two optimal sequences.  $\square$

**Remark.** The properties of the  $VPS_{\max}$  problem can also be applied to constrain the solution of the CTV problems and obtain a solution closer to the optimal one.

**Example 2.** Let us look at sequence  $C = \{9, 8, 6, 4, 2, 1, 3, 5, 7\}$  and  $C' = \{9, 8, 5, 3, 2, 1, 4, 6, 7\}$ . It is observed that both sequences are V-shaped [4] and the three largest elements are placed in the optimal positions (i.e. the largest element has to be placed in position 1 while the second and third largest elements should be placed in position 2 and  $n$ ) [3]. Then we could not determine which one is better in the sense of smaller variance based on existing theory. However, using Theorem 1, we immediately see that  $C'$  is better with smaller variance for the CTV problems.

## Appendix

The proof of Lemma 1 is provided in this section. Let  $\alpha = 2c_1 + c_2 + c_3 + \dots + c_n$  and  $\mathbf{S}^d = \{s_k^d, 1 \leq k \leq n\}$  denote the partial sum sequence of  $\mathbf{C}^d$ . Note that

$$\begin{aligned} s_k &= \sum_{m=1}^k c_m, s_k^d = c_1 + \sum_{m=2}^k c_{n+2-m}, 2 \leq k \leq n \\ s_1 &= s_1^d = c_1, s_n = s_n^d \\ s_{n+1-k} + s_k^d &= s_k + s_{n+1-k}^d = \alpha \\ \bar{\mathbf{S}} &= c_1 + \sum_{k=2}^n \frac{n+1-k}{n} c_k, \bar{\mathbf{S}}^d = c_1 + \sum_{k=2}^n \frac{n+1-k}{n} c_{n+2-k} \\ \bar{\mathbf{S}} + \bar{\mathbf{S}}^d &= 2c_1 + \sum_{k=2}^n \frac{n+1-k}{n} c_k + \sum_{k=2}^n \frac{k-1}{n} c_k = \alpha \end{aligned}$$

On the other hand

$$\begin{aligned} s_{n+1-k} - s_{n+1-k}^d &= s_k - s_k^d = \sum_{m=2}^k c_m - \sum_{m=n+2-k}^n c_m \\ n(\bar{\mathbf{S}} - \bar{\mathbf{S}}^d) &= \sum_{k=2}^n (n+1-k)c_k - \sum_{k=2}^n (k-1)c_k = \sum_{k=2}^n (n+2-2k)c_k \\ n(\bar{\mathbf{S}}^2 - \bar{\mathbf{S}}^{d^2}) &= \alpha \sum_{k=2}^n (n+2-2k)c_k \end{aligned}$$

Therefore

$$\begin{aligned} \sum_{k=2}^{n-1} (s_k^2 - (s_k^d)^2) &= \sum_{k=2}^{n-1} \left[ (s_k^2 - (s_k^d)^2) + (s_{n+1-k}^2 - (s_{n+1-k}^d)^2) \right] / 2 \\ &= 2\alpha \sum_{k=2}^n (s_k - s_k^d) / 2 \\ &= \alpha \sum_{k=2}^n \left( \sum_{m=2}^k c_m - \sum_{m=n+2-k}^n c_m \right) \\ &= \alpha \sum_{k=2}^n (n+1-k)c_k - \alpha \sum_{k=2}^n (k-1)c_k \\ &= n(\bar{\mathbf{S}}^2 - \bar{\mathbf{S}}^{d^2}) \end{aligned}$$

We obtain

$$\begin{aligned} f(\mathbf{C}) - f(\mathbf{C}^d) &= \left( \frac{1}{n} \sum_{k=1}^n s_k^2 - \bar{\mathbf{S}}^2 \right) - \left( \frac{1}{n} \sum_{k=1}^n (s_k^d)^2 - \bar{\mathbf{S}}^d{}^2 \right) \\ &= \frac{1}{n} \left( \sum_{k=2}^{n-1} (s_k^2 - (s_k^d)^2) - n(\bar{\mathbf{S}}^2 - \bar{\mathbf{S}}^d{}^2) \right) = 0 \end{aligned}$$

## References

1. A.G.Merten and M.E.Muller, *Variance minimization in single machine sequencing problems*, Manag. Sci. 18(9), 518-528, 1972.
2. L. Schrage, *Minimizing the time-in-system variance for a finite jobset*, Manag. Sci. 21(5), 540-543, 1975.
3. N.G.Hall and W.Kubiak, *Proof of a conjecture of Schrage about the completion time variance problem*, Oper. Res. Lett. 10(8), 467-472, 1991.
4. S. Eilon and I. G. Chowdhury, *Minimizing waiting time variance in the single machine problem*, Manag. Sci. 23(6), 567-575, 1977.
5. W. Kubiak, *Completion time variance minimization on a single machine is difficult*, Oper. Res. Lett. 14(1), 49-59, 1993.
6. T.C.E. Cheng and M.Y. Kovalyov, *Batch scheduling and common due-date assignment on a single machine*, Discrete Applied Mathematic. 70, 231-245, 1996.
7. G. Viswanathkumar and G .Srinivasan, *A branch and bound algorithm to minimize completion time variance on a single processor*, Comput. Oper. Res. 30(8), 1135-1150, 2003.
8. N. Ye, X. Li, T. Farley, and X. Xu, *Job scheduling methods for reducing waiting time variance*, Comput. Oper. Res. 34(10), 3069-3083, 2007.
9. M. Maroti, B. Kusy, G. Balogh, P. Volgyesi, A. Nadas, K. Molnar, S. Dora, and A. Ledeczi, *Radio interferometric geolocation*, In Proc. of 3rd ACM International Conference on Embedded Networked Sensor Systems (SenSys), 2005.